

Summary of the Invention

Generally, the present invention provides a speech synthesis system that utilizes a pitch contour resulting in a more natural-sounding speech. The present invention modifies the predicted pitch, $b(t)$, for synthesized speech using a low frequency energy booster. The low frequency energy booster interpolates the discrete pitch values, if necessary, and increase the amount of energy of the pitch contour associated with low frequency values, such as all frequency values below 10 Hertz. The amount of energy of the pitch contour associated with low frequency values can be increased, for example, by adding band-limited noise (a carrier signal) to the pitch contour, $b(t)$, or by filtering the pitch values with an impulse response filter having a pole at the desired low frequency value. The present invention serves to add vibrato to the original pitch contour, $b(t)$, and improves the naturalness of the synthetic waveform.

A more complete understanding of the present invention, as well as further features and advantages of the present invention, will be obtained by reference to the following detailed description and drawings.

Brief Description of the Drawings

FIG. 1 is a schematic block diagram of a conventional speech synthesis system;

FIG. 2 is a schematic block diagram of a speech synthesis system in accordance with the present invention;

FIG. 3 is a frequency spectrum illustrating a certain amount of bravado that is added to the original pitch contour, $b(t)$, in accordance with the present invention; and

FIG. 4 is a flow chart describing an exemplary concatenative text-to-speech synthesis system incorporating features of the present invention.

Detailed Description of Preferred Embodiments

FIG. 2 is a schematic block diagram illustrating a speech synthesis system 200 in accordance with the present invention. The present invention is directed to a method and apparatus for synthesizing speech that utilizes an improved pitch contour resulting in a more natural-sounding speech.

As shown in FIG. 2, the speech synthesis system 200 includes the conventional speech synthesis system 100, discussed above, as well as a low frequency energy booster 220. The conventional speech synthesis system 100 may be embodied as the ETI-Eloquence 5.0, commercially available from Eloquent Technology, Inc. of Ithaca, NY, as modified herein to provide the features and functions of the present invention. As shown in FIG. 2, the conventional speech synthesis system 100 includes a pitch predictor 210 that predicts the pitch, $b(t)$, of the utterance associated with the input text, in a known manner. As previously indicated, the predicted pitch, $b(t)$, provides a pitch value specified for each syllable.

According to a feature of the present invention, the predicted pitch, $b(t)$, is modified by the low frequency energy booster 220 to interpolate the discrete pitch values and increase the amount of energy of the pitch contour associated with low frequency values, such as below 10 Hertz. The amount of energy of the pitch contour associated with low frequency values can be increased, for example, by adding band-limited noise (a carrier signal) to the pitch contour, $b(t)$. In this manner, the use of the carrier signal contributes vibrato 310 to the original pitch contour, $b(t)$, as shown in FIG. 3, and improves the naturalness of the synthetic waveform.

Thus, in one implementation, the vibrato 310 corresponds to a periodic carrier waveform, $p(t)$, added to the pitch contour, $b(t)$. Thus, the pitch frequency, $f(t)$, of the speech 230 generated by the speech synthesis system 200 can be expressed as follows:

$$f(t) = b(t) + p(t),$$

where $p(t) = a \sin(\omega t + \Phi)$;